J. Math. Anal. Appl. 363 (2010) 569–578

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

On consistency of stationary points of stochastic optimization problems in a Banach space

Pedro Terán

Escuela Universitaria de Ingeniería Técnica Industrial, Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, E-33071 Gijón, Spain

ARTICLE INFO

Article history:
Received 31 January 2009
Available online 18 September 2009
Submitted by M. Ledoux

Keywords:
Clarke generalized gradient
Dual Banach space
Komlós theorem
Stationary point
Stochastic optimization
Strong Law of Large Numbers

ABSTRACT

Recently, Balaji and Xu studied the consistency of stationary points, in the sense of the Clarke generalized gradient, for the sample average approximations to a one-stage stochastic optimization problem in a separable Banach space with separable dual. We present an alternative approach, showing that the restrictive assumptions that the dual space is separable and the Clarke generalized gradient is a (norm) upper semicontinuous and compact-valued multifunction can be dropped. For that purpose, we use two results having independent interest: a strong law of large numbers and a multivalued Komlós theorem in the dual to a separable Banach space, and a result on the weak* closedness of the expectation of a random weak* compact convex set.

1. Introduction

In their paper [1], Balaji and Xu study the approximation of the stationary points of the classical stochastic optimization problem

$$\min_{x\in\mathcal{X}} E\big[f(x,\xi)\big] \tag{1}$$

by those of its sample average approximation

$$\min_{x\in\mathcal{X}} n^{-1}\sum_{i=1}^{n} f\big(x,\xi^i\big), \tag{2}$$

where the ξ^i are i.i.d. as ξ, and f is Lipschitz. More precisely, they present the following result [1, Theorem 3].

Theorem 1.1. *Let* $\mathbf{E}$ *be a separable Banach space with separable dual, and* $\mathcal{X}$ *a compact subset of* $\mathbf{E}$*. Let* $f:\mathcal{X}\times\Omega\to\mathbf{R}$ *be such that*

(a) $Ef(x,\cdot)$ *is finite for all* $x\in\mathbf{E}$.
(b) *There exists* $\kappa\in L^1(P)$ *such that*

$$\big|f(x,\omega)-f(y,\omega)\big|\leqslant\kappa(\omega)\|x-y\|$$

for all $x,y\in\mathcal{X}$ *and almost every* $\omega\in\Omega$.
(c) $\partial_x f(\cdot,\omega)$ *is upper semicontinuous in the norm for almost every* $\omega\in\Omega$.

E-mail address: teranpedro@uniovi.es.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2009.07.062

Then, every accumulation point of a sequence of Clarke stationary points of the approximate problem (2) *is almost surely a weak Clarke stationary point of* (1).

They derive their result with the following strategy (see also the related papers [10,12]). Start with a Strong Law of Large Numbers for random sets (which replaces the ordinary SLLN for random variables used when solutions are guaranteed to be unique). Using that result, prove that a related asymptotic formula holds uniformly over $\mathcal{X}$. Finally, the latter is applied to a generalized gradient mapping, allowing to show that, for any sequence whose n-th element is a stationary point of the n-th sample approximate problem, its accumulation points are weak stationary points of the original problem.

Since, in [1], the function f is defined on a separable (maybe infinite-dimensional) Banach space $\mathbf{E}$ and the generalized gradient takes on values in the dual space, the SLLN for random sets must be applied in a dual space. But, in fact, the existing SLLNs are not well-suited for application in a dual space because they assume that the space is (norm) separable, while the dual to a separable Banach space can be non-separable. Thus, Balaji and Xu must impose the assumption that $\mathbf{E}$ *has separable dual*. Moreover, the SLLN with convergence in the Hausdorff metric is available only for random compact sets, whence Balaji and Xu make the assumption that the Clarke generalized gradient of f is *compact-valued*. And, finally, the use of the norm topology in the dual requires yet another assumption on the generalized gradient: that it is an *upper semicontinuous multifunction* with respect to the norm topology.

In this paper, we develop a Strong Law of Large Numbers valid for random weak* compact convex sets in the dual to any separable Banach space, regardless of the dual being norm separable or not. That leads to showing that all the three assumptions mentioned in the last paragraph can be dropped, so Balaji and Xu's result on consistency of stationary points actually holds under much weaker assumptions on both $\mathbf{E}$ and f.

2. Preliminaries

Let $\mathbf{E}$ be a real separable Banach space, with norm $\|\cdot\|$, closed unit ball $\mathcal{B}$ and dual $\mathbf{E}^*$ with norm $\|\cdot\|_*$. The unit sphere of $\mathbf{E}$ will be denoted S. Throughout the paper, $D=\{x_k\}_{k\in\mathbf{N}}$ will denote a fixed countable dense subset of S. The unit ball of $\mathbf{E}^*$ will be denoted by $\mathcal{B}^*$.

We will denote by j the canonical embedding of $\mathbf{E}$ into its bidual $\mathbf{E}^{**}$.

Weak* closed bounded subsets of $\mathbf{E}^*$ are weak* compact and metrizable. The topology of the separable metric d^* defined by

$$d^*(f,g)=\sum_{k\in\mathbf{N}}2^{-k}\big|f(x_k)-g(x_k)\big|$$

agrees with the weak* topology on every bounded subset of $\mathbf{E}^*$. Of course, $d^*(f,g)\leqslant\|f-g\|_*$.

The space of all non-empty closed subsets of a metric space M will be called $\mathcal{F}(M)$. The space of all non-empty weak* compact convex subsets of $\mathbf{E}^*$ will be called $\mathcal{W}^*_{kc}$. For any z in the bidual $\mathbf{E}^{**}$, we define the support function $s(z,\cdot):\mathcal{W}^*_{kc}\to\mathbf{R}$, given by $s(z,A)=\sup_{f\in A}z(f)$. Then,

$$A=\bigcap_{x\in D}\big\{f\in\mathbf{E}^*\;\big|\;f(x)\leqslant s\big(j(x),A\big)\big\}.$$

Addition and product by scalars in $\mathcal{W}^*_{kc}$ are defined elementwise; since linear operations are weak* continuous, $\mathcal{W}^*_{kc}$ is closed under the extended operations. The convex hull and closed convex hull of A are denoted by $\operatorname{co}A$ and $\overline{\operatorname{co}}\,A$, respectively. Topological closure is denoted by cl; if the topology needs to be specified, we will write e.g. d^*-cl.

We denote

$$d^*(f,A)=\inf_{g\in A}d^*(f,g)$$

and

$$D^*(A,C)=\sup_{f\in A}d^*(f,A).$$

The *Hausdorff metric* associated to d^* is defined to be

$$H^*(A,C)=\max\big\{D^*(A,C),D^*(C,A)\big\}.$$

We also define

$$\|A\|_*=\sup_{f\in A}\|f\|_*=D^*\big(A,\{0\}\big)=H^*\big(A,\{0\}\big).$$

Note that D^* and H^* make sense for arbitrary non-empty subsets but, for sets which are not d^*-bounded and d^*-closed, some of the properties defining a metric are lost.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a mapping from Ω to $\mathcal{F}(\mathbf{E}^*, d^*)$. X is called a *random* (d^*-)*closed set* if it is *Effros measurable*, namely the event $\{X \cap U \neq \emptyset\}$ is $\mathcal{A}$-measurable for each open set U. Since we will be working with random closed sets in $\mathbf{E}^*$, the underlying topology will be that of d^*.

We denote by $S^1(X)$ the set of all integrable selections of X. The *expectation* of X is then defined to be

$$E_I X = \{E\xi \mid \xi \in S^1(X)\}.$$

Sometimes the expectation is defined as the closure of $E_I X$; to emphasize that it is not the case here, we adopt the notation E_I instead of E, following [8].

If $E\|X\|_* < \infty$, then X is called *integrably bounded*.

Let $(\mathcal{X}, \rho)$ be a compact metric space. A set-valued function $\mathcal{A}$ on $\mathcal{X}$ is called *upper semicontinuous* if, for every $x \in \mathcal{X}$ and every open neighbourhood U of $\mathcal{A}(x)$, there is $\delta > 0$ such that

$$\rho(x, y) < \delta \quad \Rightarrow \quad \mathcal{A}(y) \subset U.$$

Balaji and Xu [1] considered upper semicontinuity by taking the norm topology in $\mathbf{E}^*$, while we will take the topology of d^*. Thus, our notion is weaker than theirs. If the values of $\mathcal{A}$ are d^*-compact, then $\mathcal{A}$ is upper semicontinuous with respect to d^* if and only if

$$D^*\big(\mathcal{A}(y), \mathcal{A}(x)\big) \to 0 \quad \text{as } y \to x.$$

3. Laws of large numbers

In this section, we prove the laws of large numbers which are at the basis of the consistency result in Section 5. We begin with the following lemmas.

Lemma 3.1. *Let $A \in \mathcal{W}^*_{kc}$. Then, $\mathcal{F}(A)$ is H^*-compact.*

Proof. Since the weak* and d^* topologies agree on bounded sets, A is d^*-compact. By classical results in hyperspace topology (see e.g. [9]), given any compact metric space K, the space $\mathcal{F}(K)$ is compact in the Vietoris topology; and so in the Hausdorff metric also (see e.g. [8, Corollary C.6, p. 404]). □

Lemma 3.2. *Let $\lambda_i > 0$ and $A_i, C_i \in \mathcal{W}^*_{kc}$. Then,*

$$D^*\left(\sum_{i=1}^n \lambda_i A_i, \sum_{i=1}^n \lambda_i C_i\right) \leqslant \sum_{i=1}^n \lambda_i D^*(A_i, C_i).$$

Proof. The inequality

$$d^*\left(\sum_i \lambda_i f_i, \sum_i \lambda_i g_i\right) \leqslant \sum_i \lambda_i d^*(f_i, g_i)$$

is easy. Then, the extension to D^* is done similarly to that of the analogous known inequality in which d^* is replaced by the norm. □

Now we proceed to presenting the Strong Law of Large Numbers in $\mathcal{W}^*_{kc}$. To the best of our knowledge, this is the first SLLN for random sets in a Banach space which may be non-separable in the norm topology.

Theorem 3.3. *Let $\mathbf{E}^*$ be the dual to a separable Banach space. Let X be an integrably bounded random weak* closed convex set in $\mathbf{E}^*$, and let $\{X_n\}_n$ be pairwise independent and identically distributed as X. Then,*

$$H^*\left(n^{-1}\sum_{i=1}^n X_i, E_I X\right) \to 0$$

almost surely.

Proof. We denote by τ_D the weak topology in $\mathcal{W}^*_{kc}$ generated by the mappings $s(j(x), \cdot)$ for all $x \in D$. By the linearity of support functions,

$$s\left(j(x), n^{-1}\sum_{i=1}^n X_i\right) = n^{-1}\sum_{i=1}^n s\big(j(x), X_i\big);$$

also, $s(j(x), E_I X) = Es(j(x), X)$ (e.g. [8, Theorem 1.22, p. 157]). Since $Es(j(x), X) \leqslant E\|X\|_* < \infty$, the application of Etemadi's SLLN [6] to the random variables $\{s(j(x), X_n)\}_n$ for each x in the countable set D yields

$$n^{-1}\sum_{i=1}^{n} X_i \to E_I X \quad \text{a.s. in the sense of } \tau_D.$$

The proof will be accomplished if we show that $\{n^{-1}\sum_{i=1}^{n} X_i\}_n$ is relatively H^*-compact almost surely and that H^*-convergence is stronger than τ_D-convergence. Indeed, assume that were the case. Every subsequence of $\{n^{-1}\sum_{i=1}^{n} X_i\}_n$ would have an H^*-convergent further subsequence; but its H^*-limit must be its τ_D-limit $E_I X$. Since each subsequence has a further subsequence converging to $E_I X$, the whole sequence converges to $E_I X$.

To prove the first claim, note that

$$\left\| n^{-1}\sum_{i=1}^{n} X_i \right\|_* \leqslant n^{-1}\sum_{i=1}^{n} \|X_i\|_* \to E\|X\|_*$$

almost surely, by an application of the SLLN to the random variables $\{\|X_n\|_*\}_n$. There exists $n_0 \in \mathbf{N}$ such that, for all $n \geqslant n_0$,

$$\left\| n^{-1}\sum_{i=1}^{n} X_i \right\|_* \leqslant E\|X\|_* + 1.$$

Therefore, taking

$$r = \max\left\{ E\|X\|_* + 1, \max_{n<n_0} \left\| n^{-1}\sum_{i=1}^{n} X_i \right\|_* \right\},$$

we can write

$$n^{-1}\sum_{i=1}^{n} X_i \subset r\mathcal{B}^*$$

for all $n \in \mathbf{N}$. Then,

$$\left\{ n^{-1}\sum_{i=1}^{n} X_i \right\}_n \subset \mathcal{F}(r\mathcal{B}^*)$$

and $\mathcal{F}(r\mathcal{B}^*)$ is H^*-compact by Lemma 3.1.

To prove the second claim, let $\{A_n\}_n \subset \mathcal{W}^*_{kc}$ converge to A in d^*, and fix $x \in D$. Let us show that $s(j(x), A_n) \to s(j(x), A)$.

By the definition of d^*, there exists some $k \in \mathbf{N}$ such that, for all $f, g \in \mathbf{E}^*$,

$$d^*(f, g) \geqslant 2^{-k}\bigl|f(x) - g(x)\bigr|.$$

Fix $\varepsilon > 0$. For n large enough, it holds that for any $f \in A_n$ there is some $g \in A$ such that $d^*(f, g) \leqslant 2^{-k}\varepsilon$, and so $|f(x) - g(x)| \leqslant \varepsilon$. Accordingly,

$$s\bigl(j(x), A_n\bigr) - s\bigl(j(x), A\bigr) = \sup_{f\in A_n} f(x) - \sup_{g\in A} g(x) \leqslant \sup_{g\in A} g(x) + \varepsilon - \sup_{g\in A} g(x) = \varepsilon.$$

A similar argument shows that $s(j(x), A) - s(j(x), A_n) \leqslant \varepsilon$, and so

$$\bigl|s\bigl(j(x), A_n\bigr) - s\bigl(j(x), A\bigr)\bigr| \leqslant \varepsilon.$$

Since that proves $s(j(x), A_n) \to s(j(x), A)$ for any $x \in D$, we have $A_n \to A$ in the topology τ_D. $\square$

The following uniform variant will be used in the proof of Theorem 5.3.

Theorem 3.4. *Let $(\mathcal{X}, \rho)$ be a compact metric space, and let $\mathbf{E}^*$ be the dual to a separable Banach space. Let $\mathcal{A} : \mathcal{X} \times \Omega \to \mathcal{W}^*_{kc}$ be such that*

(i) *$\mathcal{A}(x, \cdot)$ is measurable for each $x \in \mathcal{X}$.*
(ii) *$\mathcal{A}(\cdot, \omega)$ is an upper semicontinuous multifunction with respect to d^*, for almost every $\omega \in \Omega$.*
(iii) *$|\sup_{x\in\mathcal{X}} \mathcal{A}(x, \cdot)| \leqslant \phi$ for some $\phi \in L^1(P)$.*

Finally, let $\{\mathcal{A}_n\}_n$ be a sequence of pairwise i.i.d. copies of $\mathcal{A}$. Then,

$$\sup_{x\in\mathcal{X}} D^*\big(\mathcal{S}_n(x), E_I\big[\mathcal{A}^r(x,\cdot)\big]\big) \to 0$$

almost surely for every $r > 0$, where

$$\mathcal{A}^r(x,\omega) = \bigcup_{\rho(y,x)\leqslant r} \mathcal{A}(y,\omega).$$

Proof. The proof is completely analogous to that of [1, Theorem 1], taking into account Lemma 3.2 and Theorem 3.3. □

Remark 3.1. A direct comparison between Theorem 3.4 and [1, Theorem 1] is not possible. First, they do not hold in the same spaces. Moreover, the class of spaces where both theorems hold is that of separable dual spaces; then, our result has weaker assumptions (pairwise independence instead of independence, upper semicontinuity with respect to d^* instead of $\|\cdot\|_*$) but a correspondingly weaker conclusion.

Thus both results are complementary, although Theorem 3.4 will be shown to be better suited for the application to convergence of stationary points.

4. Komlós theorem and closedness of the expectation

The aim of this section is to prove the following result, which will be necessary at the end of the proof of Theorem 5.3. However, all results in this section have independent significance. For related results in the setting of separable Banach spaces, see e.g. [8, Theorem 1.24, p. 158] (closedness of the Aumann expectation) and [8, Theorem 3.3, p. 126] (Komlós theorem).

Proposition 4.1. *Let $\mathbf{E}^*$ be the dual to a separable Banach space. Let X be an integrably bounded random weak* closed convex set in $\mathbf{E}^*$. Then, $E_I X$ is a non-empty weak* compact convex subset of $\mathbf{E}^*$.*

A key ingredient of the proof is a suitable version of the Komlós theorem. The result we need is a variant for the weak* topology of [2, Corollary 2.2]; the latter is obtained as an immediate particularization of the multivalued Komlós theorem [2, Theorem 2.1]. Since the proof is a routine adaptation of theirs, we state here both the multivalued and the single-valued version.

Theorem 4.2. *Let $\mathbf{E}^*$ be the dual to a separable Banach space. Let $\{X_n\}_n$ be a sequence of random weak* closed convex sets such that $\sup_n E\|X_n\|_* < \infty$. Then, there exist a subsequence $\{n'\}_n$ and an integrably bounded random weak* compact convex set X such that*

(a) *$X(\omega) \subset \overline{\mathrm{co}}\, d^*\text{-}\limsup_n X_{n'}(\omega)$ for almost every $\omega \in \Omega$.*
(b) *For any further subsequence $\{n''\}_n$,*

$$m^{-1}\sum_{n=1}^{m} X_{n''} \to X$$

almost surely in H^*.

Proof. Except for some changes, it is very similar to that of [2, Theorem 2.1] and so left to the reader. The role of the spaces X, X^* in their proof is played by $\mathbf{E}^*$, $j(\mathbf{E})$; the Mackey and weak topologies in X^* are replaced by the norm topology on $\mathbf{E}$; weakly compact sets are replaced by weak* compact sets; and the unit ball of X^* can be replaced by S (where our set D is dense).

In step 2 of the proof, scalar convergence is replaced by H^*-convergence. The analog of [2, Lemma 3.2] for H^* has been proved in Lemma 3.1. Concerning that step, observe that, by defining

$$G(\omega) = d^*\text{-cl co}\bigcup_n X_n(\omega)$$

with d^*-cl in place of cl, the intersection of $G(\omega)$ and a ball is necessarily d^*-compact, so an analog of assumption (ii) in [2, Theorem 2.1] is not needed.

In step 3, the measurability of $d^*\text{-}\limsup_n X_{n'}$ holds because d^* is separable. (Note: later, in the argument involving weak compactness in an L^1 space, weak compactness should not be replaced by compactness in any other topology; when [3, Theorem 8] is used, just note that the weak limsup is contained in the d^*-limsup.)

Apart from these changes, the essence of the proof is identical. □

Remark 4.1. Note that a weaker variant of Theorem 3.3, with independence instead of pairwise independence, can be derived from Theorem 4.2, by showing that the limit in (b) is deterministic (it is measurable with respect to the tail σ-algebra), then identifying it as $E_I X$.

Corollary 4.3. *Let $\mathbf{E}^*$ be the dual to a separable Banach space. Let $\{\xi_n\}_n$ be a sequence of Bochner integrable functions such that $\sup_n E\|\xi_n\|_* < \infty$. Then, there exist a subsequence $\{n'\}_n$ and a Bochner integrable function ξ such that*

(a) *$\xi(\omega) \in \overline{\mathrm{co}}\, d^*\text{-}\limsup_n \{\xi_{n'}(\omega)\}$ for almost every $\omega \in \Omega$.*
(b) *For any further subsequence $\{n''\}_n$,*

$$m^{-1}\sum_{n=1}^{m} \xi_{n''} \to \xi$$

almost surely in d^*.

Proof. Just take $X_n = \{\xi_n\}$ and note that Theorem 4.2(b) implies that X must be a singleton as well. The facts that $X = \{\xi\}$ is Effros measurable and integrably bounded mean that ξ is Borel measurable and Bochner integrable. □

Proof of Proposition 4.1. We begin by proving that $E_I X$ is non-empty. To that end, note that X takes on d^*-compact values. By Sion's selection theorem for compact-valued mappings [11], X admits a measurable selection ξ. But $\|\xi\|_* \leqslant \|X\|_*$ and $E\|X\|_* < \infty$, so ξ is Bochner integrable. Then $E\xi \in E_I X$.

Now, it is easy to prove that $E_I X$ is bounded and convex, from the fact that $E_I X \subset E\|X\|_* \cdot \mathcal{B}$ and the convexity of the values of X.

There remains to prove that $E_I X$ is weak* closed. Let $\{f_n\}_n \subset E_I X$ be such that f_n weak* converges to $f \in \mathbf{E}^*$, we have to prove $f \in E_I X$. By definition, $f_n = E\xi_n$ for some $\xi_n \in S^1(X)$.

Note that Corollary 4.3 applies to $\{\xi_n\}_n$, since $\sup_n E\|\xi_n\|_* \leqslant E\|X\|_*$ and X is integrably bounded. Thus, there exists a Bochner integrable function ξ which, by virtue of its conclusion (a) and the weak* closedness of the values of X, is indeed a selection of X; and a subsequence $\{\xi_{n'}\}_n$ such that

$$d^*\left(m^{-1}\sum_{n=1}^{m} \xi_{n'}, \xi\right) \to 0$$

almost surely. Since $\|m^{-1}\sum_{n=1}^{m} E\xi_{n'}\|_* \leqslant E\|X\|_*$, the sequence $\{m^{-1}\sum_{n=1}^{m} E\xi_{n'}\}_m$ is contained in the ball $E\|X\|_* \cdot \mathcal{B}$, where d^* metrizes the weak* topology. And, since the sequence $\{\|\xi_{n'}\|\}_n$ is dominated by $\|X\|_*$, the dominated convergence theorem yields now

$$m^{-1}\sum_{n=1}^{m} (E\xi_{n'})(x) \to (E\xi)(x)$$

for all $x \in \mathbf{E}$.

But $E\xi_{n'} \to f$ in the weak* topology, so for each $x \in \mathbf{E}$ we have $(E\xi_{n'})(x) \to f(x)$. By the regularity of Cesàro summability, also

$$m^{-1}\sum_{n=1}^{m} (E\xi_{n'})(x) \to f(x).$$

By the uniqueness of the weak* limit, $f = E\xi$; and, since ξ is a selection of X, we have $f \in E_I X$, proving that $E_I X$ is weak* closed. □

5. Convergence of stationary points

In this section, we apply the results above to analyze convergence of the stationary points of the sample average approximate problem (2).

Let $h : \mathbf{E} \to \mathbf{R}$ be locally Lipschitz at a point x. Its *Clarke generalized gradient* $\partial h(x) \subset \mathbf{E}^*$ is the unique element of $\mathcal{W}^*_{kc}$ such that

$$s\big(j(d), \partial h(x)\big) = \limsup_{y \to x, t \to 0^+} \frac{h(y + td) - h(y)}{t}$$

for all $d \in S$.

Now let $\mathcal{X}$ be a compact subset of $\mathbf{E}$ and $f:\mathcal{X}\times\Omega\to\mathbf{R}$. We denote by $\partial_x f:\mathcal{X}\times\Omega\to\mathcal{W}^*_{kc}$ the mapping given by $\partial_x f(y,\omega)=\partial f(\cdot,\omega)(y)$.

Let $\{f_n\}_n$ be a sequence of pairwise independent copies of f. The specific meaning of 'pairwise independent' is the one when f is regarded as a random lower semicontinuous function, see e.g. [7]. That includes, as a special case, the composition $f_n=f(\cdot,\xi^n)$ involving a deterministic function f and pairwise independent random elements ξ^n.

Set

$$\hat{f}_n(x)=n^{-1}\sum_{i=1}^n f_i(x,\cdot),$$

$$S_n(x)=n^{-1}\sum_{i=1}^n \partial_x f_i(x,\cdot),$$

$$\partial_x^r f_i(x,\cdot)=\bigcup_{y\in\mathcal{X},\,\|x-y\|<r}\partial_x f_i(y,\cdot),$$

$$S_n^r(x)=n^{-1}\sum_{i=1}^n \partial_x^r f_i(x,\cdot),$$

where $r>0$.

We denote by $\mathcal{N}_{\mathcal{X}}(x)$ the normal cone to $\mathcal{X}$ at x. A point $x^*\in\mathcal{X}$ is called a *Clarke stationary point of* (1) if

$$0\in\partial Ef(x^*,\cdot)+\mathcal{N}_{\mathcal{X}}(x^*)$$

and a *weak Clarke stationary point of* (1) if

$$0\in E_I\partial f(x^*,\cdot)+\mathcal{N}_{\mathcal{X}}(x^*),$$

see [1]. Analogously, a point $x_n\in\mathcal{X}$ is called a *Clarke stationary point of* (2) if

$$0\in\partial\hat{f}_n(x_n)+\mathcal{N}_{\mathcal{X}}(x_n).$$

We still need one further lemma which will be used in the proof of Theorem 5.3.

Lemma 5.1. *Let X, Y be random closed sets in the dual of a separable Banach space. Assume that $E_IX\neq\emptyset$ and Y is integrably bounded. Then, $D^*(E_IX,E_IY)\leqslant ED^*(X,Y)$.*

Proof. Let ξ be an arbitrary integrable selection of X. Observe that

$$d^*(E\xi,E_IY)=\inf_{\eta\in S^1(Y)}d^*(E\xi,E\eta)=\inf_{\eta\in S^1(Y)}\sum_k 2^{-k}\left|E\xi(x_k)-E\eta(x_k)\right|$$
$$\leqslant\inf_{\eta\in S^1(Y)}E\left[\sum_k 2^{-k}\left|\xi(x_k)-\eta(x_k)\right|\right]=\inf_{\eta\in S^1(Y)}Ed^*(\xi,\eta).$$

Set

$$Y_n=w^*\text{-}\mathrm{cl}\left[Y\cap\left\{f\in\mathbf{E}^*\;\middle|\;d^*(\xi,f)\leqslant d^*(\xi,Y)+n^{-1}\right\}\right].$$

Since Y is integrably bounded, it takes on bounded values almost surely, so Y_n is weak* closed and norm bounded, whence it is weak* compact, equivalently d^*-compact. Now we use Sion's selection theorem to obtain a measurable selection $\eta_n\in Y_n$. Since Y_n, like Y, is integrably bounded, the η_n are necessarily in $S^1(Y_n)$ and so in $S^1(Y)$.

By construction,

$$d^*(\xi,\eta_n)\leqslant d^*(\xi,Y)+n^{-1},$$

so

$$d^*(E\xi,E_IY)=\inf_{\eta\in S^1(Y)}Ed^*(\xi,\eta)\leqslant\inf_n Ed^*(\xi,\eta_n)\leqslant Ed^*(\xi,Y)\leqslant ED^*(X,Y),$$

whence

$$D^*(E_IX,E_IY)=\sup_{\xi\in S^1(X)}d^*(E\xi,E_IY)\leqslant ED^*(X,Y).\qquad\square$$

In order to apply the uniform law of large numbers to the Clarke generalized gradient, we need to study its upper semicontinuity properties. According to Edalat [5], who cites a private communication from F.H. Clarke, it is unknown whether the Clarke generalized gradient is an upper semicontinuous multifunction with respect to the weak* topology. The fact that d^* is a metric simplifies the task of establishing such a property.

Proposition 5.2. *Let* $\mathbf{E}$ *be a separable Banach space. Let* $f : \mathbf{E} \to \mathbf{R}$ *be a locally Lipschitz function. Then,* ∂f *is upper semicontinuous with respect to* d^*.

Proof. Let $x \in \mathbf{E}$. The values of ∂f are weak* compact, in particular norm bounded, sets. Since d^* agrees with the weak* topology on bounded sets, ∂f takes on d^*-compact values. Consequently, to prove upper semicontinuity it suffices to check that $D^*(\partial f(y), \partial f(x)) \to 0$ as $y \to x$. For any $r > 0$, denote $\partial^r f(x) = \bigcup_{\|y-x\| \leqslant r} \partial f(y)$. We will be done by proving that $D^*(d^*\text{-}\operatorname{cl} \partial^r f(x), \partial f(x)) \to 0$ as $r \searrow 0$.

Let us prove $\partial f(x) = \bigcap_{r>0} d^*\text{-}\operatorname{cl} \partial^r f(x)$. By [4, Proposition 2.1.5(c)], $\partial f(x) = \bigcap_{r>0} \partial^r f(x)$, so one inclusion is clear. For the converse, let $g \in \bigcap_{r>0} d^*\text{-}\operatorname{cl} \partial^r f(x)$, then

$$d^*\big(g, \partial f(x)\big) = d^*\left(g, \bigcap_{r>0} \partial^r f(x)\right) = \sup_{r>0} d^*\big(g, \partial^r f(x)\big)$$

$$= \sup_{r>0} d^*\big(g, d^*\text{-}\operatorname{cl} \partial^r f(x)\big) = d^*\left(g, \bigcap_{r>0} d^*\text{-}\operatorname{cl} \partial^r f(x)\right) = 0.$$

Since $\partial f(x)$ is d^*-closed, $g \in \partial f(x)$.

Let k be a local Lipschitz constant for f, valid for y such that $\|y - x\| \leqslant r_0$ for some $r_0 > 0$. It follows easily from the definition of ∂f that $\|\partial f(y)\|_* \leqslant k$ for those y. Thus, $\|\partial^{r_0} f(x)\| \leqslant k$ and, from the Banach–Alaoglu theorem, $\partial^{r_0} f(x)$ is relatively weak* (and d^*) compact.

For every $r < r_0$, we have $d^*\text{-}\operatorname{cl} \partial^r f(x) \in \mathcal{F}(d^*\text{-}\operatorname{cl} \partial^{r_0} f(x))$. By Lemma 3.1, the latter family is H^*-compact, whence $d^*\text{-}\operatorname{cl} \partial^r f(x)$ converges in H^* along some sequence $r_n \searrow 0$. The H^*-limit must coincide with the limit inferior of the sequence, which, by the identity $\partial f(x) = \bigcap_{r>0} d^*\text{-}\operatorname{cl} \partial^r f(x)$, is $\partial f(x)$.

Therefore, $d^*\text{-}\operatorname{cl} \partial^{r_n} f(x) \to \partial f(x)$ in H^* and, taking into account the monotonicity,

$$D^*\big(d^*\text{-}\operatorname{cl} \partial^r f(x), \partial f(x)\big) \to 0$$

as $r \searrow 0$, as wished. $\square$

We are finally ready to study the consistency properties of Clarke stationary points. There are some differences with Theorem 1.1, besides dropping three assumptions as discussed in the Introduction. First, condition (a) is formally weaker than condition (a) in Theorem 1.1, although they are in fact equivalent under condition (b). Second, the independence assumption is relaxed to pairwise independence. Last, and most important, note the different place for 'almost surely' in the conclusion. While we prove that for almost every $\omega \in \Omega$ it holds that every accumulation point x of Clarke stationary points $x_{n'}(\omega) \to x$ has a certain property, what was stated then was that every sequence $x_n : \Omega \to \mathbf{E}$, such that each $x_n(\omega)$ is a Clarke stationary point and $x_{n'} \to x$ almost surely, has almost every $x(\omega)$ with that property. In our statement, the subsequence may be different for each ω. That has a practical bearing, since only one ω is observed in reality, so their formulation leaves out whether the subsequence observed to converge for that ω would admit a further subsequence valid for all ω, as needed to apply their theorem, or not.

Theorem 5.3. *Let* $\mathbf{E}$ *be a separable Banach space and* $\mathcal{X}$ *a compact subset of* $\mathbf{E}$. *Let* $f : \mathcal{X} \times \Omega \to \mathbf{R}$ *be such that*

(a) $Ef(x, \cdot)$ *is finite for some* $x \in \mathbf{E}$.
(b) *There exists* $\kappa \in L^1(P)$ *such that*

$$\big|f(x, \omega) - f(y, \omega)\big| \leqslant \kappa(\omega)\|x - y\|$$

for all $x, y \in \mathcal{X}$ *and almost every* $\omega \in \Omega$.

Let $\{f_n\}_n$ *be a sequence of pairwise independent copies of* f. *Then, almost surely, every accumulation point of a sequence of Clarke stationary points of the approximate problem*

$$\min_{x \in \mathcal{X}} n^{-1} \sum_{i=1}^{n} f_i(x, \cdot)$$

is a weak Clarke stationary point of the problem

$$\min_{x \in \mathcal{X}} E\big[f(x, \cdot)\big].$$

Proof. Let x_n be a stationary point of the n-th sample approximate problem, namely

$$0 \in \partial \hat{f}_n(x_n) + \mathcal{N}_{\mathcal{X}}(x_n),$$

and let x^* be an accumulation point of $\{x_n\}_n$. For simplicity of notation, we assume that actually $x_n \to x^*$.

Fix $r > 0$. We have

$$d^*\big(0, E_I \partial_x f(x^*, \cdot) + \mathcal{N}_{\mathcal{X}}(x^*)\big) \leqslant d^*\big(0, \partial \hat{f}_n(x_n) + \mathcal{N}_{\mathcal{X}}(x_n)\big) + D^*\big(\partial \hat{f}_n(x_n) + \mathcal{N}_{\mathcal{X}}(x_n), E_I \partial_x f(x^*, \cdot) + \mathcal{N}_{\mathcal{X}}(x^*)\big).$$

The first term in the right-hand side is 0. Note that, eventually, $\mathcal{N}_{\mathcal{X}}(x_n) \subset \mathcal{N}_{\mathcal{X}}(x^*)$ and also, by [1, Proposition 1], $\hat{f}_n(x_n) \subset S_n(x_n)$. Hence,

$$\begin{aligned} d^*\big(0, E_I \partial_x f(x^*, \cdot) + \mathcal{N}_{\mathcal{X}}(x^*)\big) &\leqslant D^*\big(S_n(x_n) + \mathcal{N}_{\mathcal{X}}(x^*), E_I \partial_x f(x^*, \cdot) + \mathcal{N}_{\mathcal{X}}(x^*)\big) \\ &\leqslant D^*\big(S_n(x_n), E_I \partial_x f(x^*, \cdot)\big) \\ &\leqslant D^*\big(S_n(x_n), S_n(x^*)\big) + D^*\big(S_n(x^*), E_I \partial_x^r f(x^*, \cdot)\big) + D^*\big(E_I \partial_x^r f(x^*, \cdot), E_I \partial_x f(x^*, \cdot)\big) \\ &=: (I) + (II) + (III). \end{aligned}$$

Since $\|x_n - x^*\| < r$ for all sufficiently large n, eventually $S_n(x_n) \subset S_n^r(x^*)$ and so, using Lemma 5.1 and Etemadi's SLLN for random variables, we obtain the almost sure bound

$$\begin{aligned} (I) &\leqslant D^*\big(S_n^r(x^*), S_n(x^*)\big) = D^*\left(n^{-1} \sum_{i=1}^n \partial_x^r f_i(x^*, \cdot), n^{-1} \sum_{i=1}^n \partial_x f_i(x^*, \cdot)\right) \\ &\leqslant n^{-1} \sum_{i=1}^n D^*\big(\partial_x^r f_i(x^*, \cdot), \partial_x f_i(x^*, \cdot)\big) \leqslant E D^*\big(\partial_x^r f(x^*, \cdot), \partial_x f(x^*, \cdot)\big) + r \\ &=: (IV) + r \end{aligned}$$

for n large enough. By Proposition 5.2 and Theorem 3.4, $(II) \to 0$ as $n \to \infty$; and an application of Lemma 5.1 yields $(III) \leqslant (IV)$. Therefore,

$$d^*\big(0, E_I \partial_x f(x^*, \cdot) + \mathcal{N}_{\mathcal{X}}(x^*)\big) \leqslant 2 \cdot (IV) + r,$$

for any arbitrary $r > 0$. By the monotone convergence theorem, to show that $(IV) \to 0$ as $r \to 0^+$, we just need to check that

$$D^*\big(\partial_x^r f_i(x^*, \omega), \partial_x f_i(x^*, \omega)\big) \to 0$$

for each $\omega \in \Omega$. But that follows from Proposition 5.2.

So far it has been proved that $d^*(0, E_I \partial_x f(x^*, \cdot) + \mathcal{N}_{\mathcal{X}}(x^*)) = 0$, whence

$$0 \in d^*\text{-}\operatorname{cl}\big[E_I \partial_x f(x^*, \cdot) + \mathcal{N}_{\mathcal{X}}(x^*)\big].$$

By Proposition 4.1, $E_I \partial_x f(x^*, \cdot)$ is d^*-compact, and we prove routinely that $\mathcal{N}_{\mathcal{X}}(x^*)$ is d^*-closed. As a consequence, their sum is d^*-closed, proving that x^* is indeed a weak Clarke stationary point. □

6. Concluding remarks

Since the properties of ∂f which are crucial to obtaining the results in this paper are related with local boundedness and the sum rule, the results can be replicated for most convex-valued generalized gradients.

Also observe that the proof of Theorem 5.3 works as well when the x_n are merely *weak* stationary points of (2).

Acknowledgment

Research supported by Spain's *Ministerio de Ciencia e Innovación* (MTM2008-01519) and Aragón's *Gobierno* (PM2007-034).

References

[1] R. Balaji, H. Xu, Approximating stationary points of stochastic optimization problems in Banach space, J. Math. Anal. Appl. 347 (2008) 333–343.
[2] E. Balder, C. Hess, Two generalizations of Komlós' theorem with lower closure-type applications, J. Convex Anal. 3 (1996) 25–44, available online at http://www.heldermann-verlag.de/jca/jca03/jca03002.pdf.
[3] A. Amrani, C. Castaing, M. Valadier, Méthodes de troncature appliquées à des problèmes de convergence faible ou forte dans L^1, Arch. Ration. Mech. Anal. 117 (1992) 167–191.
[4] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.

[5] A. Edalat, A continuous derivative for real-valued functions, in: S.B. Cooper, B. Lowe, A. Sorbi (Eds.), Computing and Logic in the Real World, in: Lecture Notes in Comput. Sci., vol. 4497, Springer, Berlin, 2007, pp. 248–257.
[6] N. Etemadi, An elementary proof of the strong law of large numbers, Z. Wahrsch. Verw. Gebiete 55 (1981) 119–122.
[7] L.A. Korf, R.J.B. Wets, Random lsc functions: An ergodic theorem, Math. Oper. Res. 26 (2001) 421–445, available online at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.79.3734&rep=rep1&type=pdf.
[8] I. Molchanov, Theory of Random Sets, Springer, London, 2005.
[9] G. Matheron, Random Sets and Integral Geometry, Wiley, New York, 1975.
[10] A. Shapiro, H. Xu, Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions, J. Math. Anal. Appl. 325 (2007) 1290–1299.
[11] M. Sion, On uniformization of sets in topological spaces, Trans. Amer. Math. Soc. 96 (1960) 237–244.
[12] P. Terán, On a uniform law of large numbers for random sets and subdifferentials of random functions, Statist. Probab. Lett. 78 (2008) 42–49.